Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 184 – 193

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Improving the Performance of a Proxy Cache Using Tree Augmented Naive Bayes Classifier

Julian Benadit.P[a,*], Sagayaraj Francis.F[a], Muruganantham.U[b]

[a]*Department of CSE, Pondicherry Engineering College, Pondicherry, 605014, INDIA*
[b]*Department of CSE, Dr.SJSPMCET, Pondicherry University, Pondicherry, 605502, INDIA*

Abstract

In this paper, we attempt to improve the performance of Web proxy cache replacement policies such as LRU and GDSF by adapting a semi naïve Bayesian learning technique. In the first part, Tree Augmented Naive Bayes classifier (TANB) to classify the web log data and predict the classes of web objects to be revisited again future or not. In the second part, a Tree Augmented Naïve Bayes classifier is incorporated with proxy caching policies to form novel approaches known as TANB-LRU and TANB-GDSF. This proposed approach improves the performances of LRU and GDSF in terms of hit and byte hit ratio respectively.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Proxy Caching;Cache replacement;Classification;Tree Augmented Naive Bayes classifier

1. Introduction

As the World Wide Web and users are explicating at a very rapid rate, the performance of World Wide Web systems becomes rapidly high. Web caching and prefetching is the one of the best methods for improving the performance of the proxy server. The basic idea in web caching is to retain the most popular web log data in a proxy cache, such that the performance of web proxy cache would have improved, when it is accessed from the user. The

* Corresponding author. Tel.:+ 919629321243.
E-mail address: benaditjulian@gmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.010

main idea behind the web caching concept is a web cache replacement, algorithm and its key parameters in the algorithms.

To ameliorate the functioning of a proxy cache, a lot of research work has been done in their cache replacement policies.Table 1. Presents a summary of the existing Cache Replacement Policies(CRP)[6,7].Most of these replacement algorithms consider only certain key factors and assign a key value based on the priority for each web document which stored in the cache. But, it is difficult to have a better cache replacement policy that performs well in all situations, because each replacement policy has a different key parameter to optimize web resources. Moreover, various elements can act upon the cache replacement policy to receive a better replacement decision and it is not an easy task for because one parameter is more significant than the other one. Due to this restriction, there is a need for an active method which intelligently manages the web proxy cache by satisfying the objectives of web caching requirement. So this platform promotes, for the integration of the machine learning methods in the web caching replacement algorithms.

In our previous surveys, the intelligent techniques have been applied in web caching algorithm. By extensive use of this prediction pattern, the caching algorithms become more efficacious and adapted for the use of web cache environment, other than the classic web caching algorithms which are already in practice. Moreover, there are multiple users who are fond of the access of web cache algorithms, but it is very necessary to prove a prediction model, which are upgraded often so that web objects can be revisited in the future on a proper standard. In this paper, we proposed the Tree Augmented Naïve Bayes classifier (TANB)[7] to train the web object based on the recurrent sliding window method for their systematic classification, so that the web objects can be predicted in the future or not. We then formulated the semi- naive Bayesian learning method called TANB to classify the web objects and then it is incorporated with traditional caching algorithm called TANB-LRU and TANB-GDSF to improve its web caching performance on a better channel.

Table 1. Cache Replacement Policies.

CRP	Key Parameter	Cache Replacement Technique
LFU	Number of References.	The least frequently accessed first.
LRU	Time Since last access.	The least recently accessed first.
GDS	Document Size S_d . Document cost C_d . An Inflation Value L.	Least value first according to value $p_i = C_d / S_d + L$.
GDSF	Document Size S_d . Document cost C_d . Number of non-aged references f_d time since last access. Temporal correlation measure β . An Inflation Value L.	Least value first according to value $p_i = (C_d \times F_d / S_d)^{\beta} + L$.

The structures of this paper with the features are processed below. In section 2, we present the related work with an Existing machine learning methods towards web caching policies. The Section 3, proposes the brief introduction of Tree Augmented Naïve Bayes classifier model. In Section 4 we introduce the proposed novel web proxy caching approach integrates with the cache replacement algorithm. Experimental results and Performance Evaluations are presented in Section 5 and Section 6. The Section 7 concludes our paper with revealed results.

2. Related Work

Web Caching holds an important role in improving the performance of the web proxy cache. The basic idea of web caching is known for its Replacement Policy, which calculates the most popular web object by storing it in the proxy cache, hence that the popular web documents can be put back with the unpopular ones. Most of the common